Addressing Artificial Intelligence Bias in Retinal Diagnostics

Year: 2021 | Citations: 112 | Authors: P. Burlina, Neil J. Joshi, W. Paul, Katia D. Pacheco, N. Bressler

Abstract

Purpose This study evaluated generative methods to potentially mitigate artificial intelligence (AI) bias when diagnosing diabetic retinopathy (DR) resulting from training data imbalance or domain generalization, which occurs when deep learning systems (DLSs) face concepts at test/inference time they were not initially trained on. Methods The public domain Kaggle EyePACS dataset (88,692 fundi and 44,346 individuals, originally diverse for ethnicity) was modified by adding clinician-annotated labels and constructing an artificial scenario of data imbalance and domain generalization by disallowing training (but not testing) exemplars for images of retinas with DR warranting referral (DR-referable) from darker-skin individuals, who presumably have greater concentration of melanin within uveal melanocytes, on average, contributing to retinal image pigmentation. A traditional/baseline diagnostic DLS was compared against new DLSs that would use training data augmented via generative models for debiasing. Results Accuracy (95% confidence intervals [CIs]) of the baseline diagnostics DLS for fundus images of lighter-skin individuals was 73.0% (66.9% to 79.2%) versus darker-skin of 60.5% (53.5% to 67.3%), demonstrating bias/disparity (delta = 12.5%; Welch t-test t = 2.670, P = 0.008) in AI performance across protected subpopulations. Using novel generative methods for addressing missing subpopulation training data (DR-referable darker-skin) achieved instead accuracy, for lighter-skin, of 72.0% (65.8% to 78.2%), and for darker-skin, of 71.5% (65.2% to 77.8%), demonstrating closer parity (delta = 0.5%) in accuracy across subpopulations (Welch t-test t = 0.111, P = 0.912). Conclusions Findings illustrate how data imbalance and domain generalization can lead to disparity of accuracy across subpopulations, and show that novel generative methods of synthetic fundus images may play a role for debiasing AI. Translational Relevance New AI methods have possible applications to address potential AI bias in DR diagnostics from fundus pigmentation, and potentially other ophthalmic DLSs too.